ON THE SPIRAL STRUCTURE OF DISK GALAXIES, II. OUTLINE OF A THEORY OF DENSITY WAVES*

BY C. C. LIN† AND FRANK H. SHU

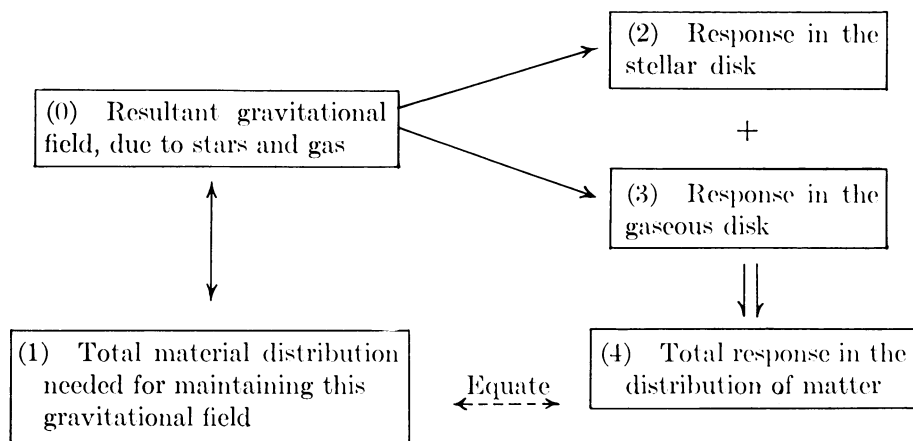INSTITUTE FOR ADVANCED STUDY, PRINCETON, NEW JERSEY,
AND HARVARD COLLEGE OBSERVATORY

1. *Introduction.*—In a previous paper[1] (hereafter referred to as Paper I or the elementary theory), we discussed a theory in which the *spiral patterns* observed in disklike galaxies are described in terms of density waves. Specifically, we stated our ideas in the form of a hypothesis of the quasi-stationary spiral structure of the stars (hereafter referred to as the QSSS hypothesis).[2] As already emphasized in that paper, the primary problem we wish to resolve is the persistence of the spiral *pattern* or the grand design over the whole disk, on the scale of 10 kpc. In such a problem, hydromagnetic forces are estimated to be negligible, to a first approximation. On the other hand, when we consider the structure of an individual spiral arm, on the scale of 1 kpc, the effect of the magnetic field must be considered.

The purpose of this note is to outline a gravitational theory that provides a full dynamical basis of the QSSS hypothesis. In particular, we wish to determine, *quantitatively*, the relative importance of the gas and the stars in providing the spiral gravitational field. Although the stars would obviously provide the predominant part of the symmetrical gravitational field, their considerable dispersive motion tends to smooth out any spiral distribution that might be present. This reduction of the effectiveness of the stars in forming a spiral structure has now been calculated on the basis of the principles of stellar dynamics (cf. §4 and Fig. 1; see also Remark 3, p. 652, Paper I). In view of this smoothing effect, we shall find that the *spiral* gravitational field of the gas is *not* negligible, although the stars still play the more important role.

2. *Self-Sustained Density Waves.*—We consider a small deviation from a symmetrical disk of stars and gas, and attempt to demonstrate that density waves of a spiral structure will be self-sustained at a small but finite amplitude. Suppose that such a wave were maintained; then there must be an associated gravitational field of a generally spiral form. We shall start with this field and carry out the analysis as indicated in the following diagram:

```
┌─────────────────────────────────┐                    ┌──────────────────────────┐
│ (0)  Resultant  gravitational   │ ──────────────────>│ (2)  Response in the     │
│ field, due to stars and gas     │                    │ stellar disk             │
└─────────────────────────────────┘                    └──────────────────────────┘
          ↑                        \                              +
          │                         \                  ┌──────────────────────────┐
          │                          \────────────────>│ (3)  Response in the     │
          │                                            │ gaseous disk             │
          ↓                                            └──────────────────────────┘
                                                                   ⇓
┌─────────────────────────────────┐                    ┌──────────────────────────┐
│ (1)  Total material distribution│      Equate        │ (4)  Total response in the│
│  needed for maintaining this    │   <- - - - - ->    │ distribution of matter    │
│ gravitational field             │                    │                          │
└─────────────────────────────────┘                    └──────────────────────────┘
```

Under the influence of the resultant spiral-like gravitational field (0) [which will eventually be considered as due to a total material distribution (1) consisting partly of gas and partly of stars], the laws of stellar dynamics would dictate a certain redistribution of the stars (2), and the laws of gas dynamics would dictate a certain redistribution of the gas (3). The sum of these two distributions yields a total distribution of matter (4) that must be equal to the density distribution (1) which we needed to give rise to the gravitational field (0) according to Poisson's equation.

This last condition is the equation to be solved for the unspecified functions and parameters that occur in the problem.

The analysis has so far been carried out only for the linear theory. Although the nonlinear effects may be expected to tend to prefer trailing patterns, we shall see that, already in the linear theory, trailing spiral waves will be found to be amplified. It may then be expected that this wave will grow to a small but finite amplitude and then attain an equilibrium through the nonlinear effects (see §6 below).

3. *The Asymptotic Theory.*—To carry out the analysis, we adopt the natural cylindrical coordinate system (ϖ, θ, z) such that the galactic disk is in the plane $z = 0$, with its center at the origin. In the linear theory, the resultant gravitational potential may be assumed to be given by a superposition of spiral modes, and the response to these individual modes may be treated separately. Let the potential of each of these modes be given by the real part of

$$\mathcal{V}_1(\varpi, \theta, t) = A(\varpi) \exp \left\{ i[\omega t - m\theta + \Phi(\varpi)] \right\}, \tag{3.1}$$

where $A(\varpi)$ and $\Phi(\varpi)$ are real functions of the axial distance, ω is a complex parameter, and m is a *positive* integer (or zero in the case of rings). The function $A(\varpi)$ is slowly varying with ϖ, whereas the function $\Phi(\varpi)$ is of the form $\epsilon^{-1}f(\varpi)$, where $f(\varpi)$ is slowly varying (and monotonic), and ϵ is a small parameter of the order of the angle of inclination of the spiral arms. Indeed, the function (3.1) clearly has a spiral structure described by the family of curves $m(\theta - \theta_0) = \Phi(\varpi) - \Phi(\varpi_0)$. The angle of inclination i of any one of these curves is given by

$$(k\varpi)^{-1} = (1/m) \tan i,$$

where $k(\varpi) = \Phi'(\varpi)$. We note that $|k(\varpi)|$ is essentially the wave number in the

radial direction. Thus, a natural approach is to adopt an asymptotic solution based on a rapidly varying phase angle, with the aid of the small parameter ϵ mentioned above. For example, if the basic symmetrical state is described by the distribution function $\Psi = \Psi_0\ (\varpi, c_{\tilde{\omega}}, c_\theta)$, where $c_{\tilde{\omega}}$ and c_θ are the peculiar velocity components of the stars relative to a state of circular motion $\Theta_c = \varpi\Omega(\varpi)$, then the distribution function in the disturbed state is given by the real part of

$$\Psi = \Psi_0(1 + \psi) \tag{3.2}$$

with

$$\psi = \varphi(\varpi, c_{\tilde{\omega}}, c_\theta) \exp\left\{i[\omega t - m\theta + \Phi(\varpi)]\right\}, \tag{3.3}$$

and

$$\varphi = \sum_{n=0}^{\infty} \epsilon^n \varphi^{(n)}(\varpi, c, c_\theta). \tag{3.4}$$

4. *Neutral Waves.*—The method used by the present writers in the elementary theory (Paper I) is essentially a special case of the scheme outlined above. Only gas and stars with zero dispersion were considered. It is expected that the high-dispersion stars would not participate in the spiral pattern in full, and that therefore the "effective" mass density of the stars must be smaller than its actual value by a suitable factor. Indeed, by following the scheme outlined in section 3 to the initial approximation, we can work out this reduction factor explicitly. If we assume that, in the basic symmetrical state, the dispersion velocities follow the Schwarzschild law of distribution (with a magnitude of dispersion sufficiently large to prevent local gravitational collapse), *neutral* density waves of the spiral form can be found. The radial wave number of such waves is given by

$$|k(\varpi)| = \frac{\kappa^2 - (\omega - m\Omega)^2}{2\pi G[\sigma_0 + \sigma_*\mathfrak{F}_\nu(x)]}, \tag{4.1}$$

for the range of values of ϖ such that

$$\kappa^2 - (\omega - m\Omega)^2 > 0. \tag{4.2}$$

In the above formulae, $\kappa(\varpi)$ is the epicyclic frequency corresponding to the circular velocity $\varpi\Omega(\varpi)$, $\sigma_0(\varpi)$ is the basic density distribution of the gas, $\sigma_*(\varpi)$ is the basic density distribution of the stars, and $\mathfrak{F}_\nu(x)$ is the *reduction factor* defined as follows:

$$\mathfrak{F}_\nu(x) = \frac{1 - \nu^2}{x}\left\{1 - \frac{\nu\pi}{\sin\nu\pi}\cdot\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-x(1 + \cos s)}\cos\nu s\, ds\right\}, \tag{4.3}$$

where $\nu = (\omega - m\Omega)/\kappa$, $x = k^2 \langle c_{\tilde{\omega}}^2 \rangle/\kappa^2$, and $\langle c_\omega^2 \rangle$ is the mean square value of the radial component of the dispersion velocity of the stars in the basic symmetrical state.

Equation (4.1) may be rewritten into the pair of relations

$$\frac{\sigma_p}{\sigma_*} = \frac{|k|}{k_*}(1 - \nu^2), \tag{4.4a}$$

$$\frac{\sigma_p}{\sigma_*} = \frac{\sigma_0}{\sigma_*} + \mathfrak{F}_\nu(x), \tag{4.4b}$$

where $k_* = \kappa^2/2\pi G\sigma_*$. We may describe σ_p/σ_* as the total "mobile component" (measured in terms of the stellar mass) which determines, according to (4.4a), the wave number $|k|$ for a given frequency ν. It consists, according to (4.4b), of a gaseous component and a stellar component.

Equation (4.1) is one of the central relationships in the present theory. By setting $x = 0$, it can be readily compared with equation (12) in Paper I. In general, $\mathfrak{F}_\nu(x)$ is real for real values of ν and x and its values are plotted in Figure 1 against ν^2 for a series of values of x.

5. *Comments on the Neutral Waves.*—(a) We note that equation (4.1) holds only for a range of values of ϖ such that (4.2) is satisfied. We shall refer to this range as the *principal part of the spiral pattern*. If we now use the Schmidt model of 1956 (for which the sun is located at 8.2 kpc) and if we adopt the value 20 km/sec-kpc as the angular velocity of the pattern $\Omega_p = \omega/m$ ($m = 2$), we find that the principal part of the spiral pattern extends over the range 2 kpc $< \varpi <$ 12 kpc. The inner radius is somewhat uncertain; the outer radius is well beyond the solar neighborhood. For other integral values of m, one finds a very limited principal part.

(b) We note that equation (4.1) does not give any indication of the sign of $k(\varpi)$. Thus, up to the present approximation, there is no way to differentiate between leading waves ($k > 0$) and trailing waves ($k < 0$). This is indeed expected from general considerations of symmetry. Trailing waves will be found to be preferred in the next approximation, which includes terms of the order of

$$\epsilon = (1/m)\tan i.$$

(c) By considering the special case $\nu = 0$ in equation (4.1), one obtains the marginal condition for neutral oscillations at a given location. For a given value of σ_0/σ_*, there is a relationship between the wavelength $2\pi/|k|$ and the velocity dispersion $\langle c_\varpi^2 \rangle$. However, realistic results can be obtained only if we apply a similar reduction factor to the gas. The result of such a calculation is shown in Figure 2, where we have taken the mean square turbulent motion to be 1/10 of $\langle c_\varpi^2 \rangle$ for the stars.[3] Conditions above the marginal curve correspond to neutral oscillations, those below to instability.

The marginal curve for zero gas content has been obtained earlier by Toomre[4] and Kalnajs.[5] Toomre also discussed the effect of the presence of the gas, but he did not examine the effect of its turbulent motion.

(d) With the help of Figure 1, one may easily determine the values of the parameters corresponding to any neutral wave. Consider σ_*, σ_0, κ^2, $\langle c_\varpi^2 \rangle$ (and hence k_*) as known. Then for a given wavelength, i.e., a chosen value of $|k|/k_*$, there is an associated value of x. If we now plot a straight line to represent (4.4a) for this value of $|k|/k_*$ (shown dashed in Fig. 1 for $k_* = 0.2|k|$) and associate it with the corresponding curve for $\mathfrak{F}_\nu(x)$ for $x = 7$ (see ref. 6), the value of ν at which the two curves separate by σ_0/σ_* is the proper value, and the ordinates on the two curves give, respectively, the total mobile component and the part due to the stellar disk. For example, the neutral wave with $\nu = 0.7$ requires a total mobile component of 11 per cent of the stellar mass, with 9 per cent from the stars and 2 per cent from the gas.
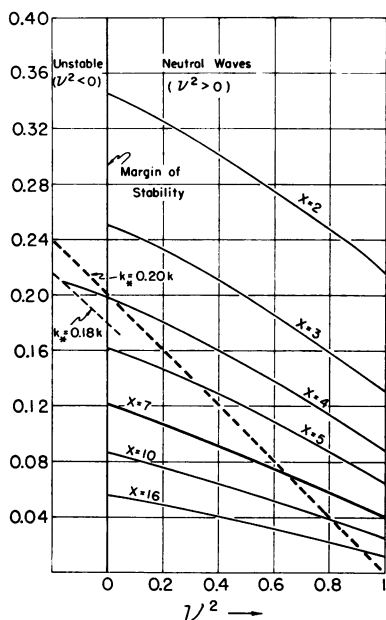
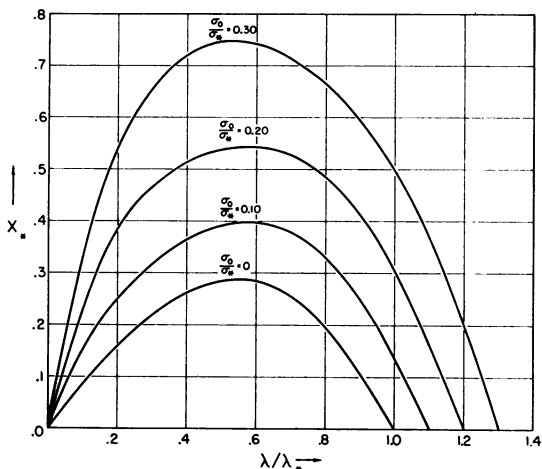FIG. 1.—Diagram showing the reduction factor $\mathfrak{F}_\nu(x)$ and the relationship (4.4).

FIG. 2.—Margin of stability when there is gas in turbulent motion. ($\lambda = 2\pi/|k|$, $\lambda_* = 2\pi/k_*$, $x_* = k_*^2 \langle c_{\tilde{\omega}}^2 \rangle / \kappa^2$.)

In particular, the conditions of marginal stability are given by the real axis $\nu^2 = 0$. Unstable conditions correspond to $\nu^2 < 0$.

6. *Amplification of Trailing Waves.*—All the above results are stated for the initial approximation in the series (3.4). In the next approximation, in which we include terms of the order of the angle of inclination of the spiral arms, we shall find indications of *instability*. This small amplification rate is then expected to be balanced by the nonlinear effects to produce neutral waves at amplitudes of the order of $\sqrt{\epsilon}$, where ϵ is of the order of $(1/m) \tan i$.

An indication of instability may already be found in the fore-aft asymmetrical deviation from the Schwarzschild distribution in the axially symmetrical basic state, even if the spiral disturbance (3.3) is calculated only to the initial approximation. But the more important effect is found only if we carry out the calculations to the order of ϵ in equation (3.4). *Amplification of trailing waves* is then found for a range of values of the radial distance for which

$$-1 < m(\Omega_p - \Omega)/\kappa < 0. \tag{6.1}$$

For the Schmidt model, with $\Omega = 20$ km/sec-kpc, this yields the range 2 kpc < $\tilde{\omega}$ < 10 kpc, which extends well beyond the solar neighborhood. The less massive part still farther out from the galactic center is expected to be driven into a trailing pattern by the more massive interior part. These self-excited trailing spiral waves may also be expected to produce a small barlike singularity near the center.

The amplification index $-\omega_i$ is given by

$$\frac{\omega_i}{\kappa} \text{ sign } (k) = \frac{\langle c_{\tilde{\omega}}^2 \rangle}{\kappa \tilde{\omega}} \cdot f(\zeta) \cdot D \tag{6.2}$$

to a first approximation for large values of x, where $\zeta = m(\Omega_p - \Omega)/\kappa$,

$$D = -d \log \langle c_{\bar{\omega}}^2 \rangle / d \log \bar{\omega} + \frac{1}{2},$$

and

$$f(\zeta) = \sqrt{\frac{2}{\pi}} \frac{\sin^2 \zeta}{\sin \zeta \cos \zeta - \zeta} > 0 \qquad (\text{for } -\pi < \zeta < 0). \tag{6.3}$$

7. *Concluding Remarks.*—Observations of the phenomena over a galactic disk show transient features as well as features which appear to be quasi-permanent. It is not difficult to account for transient features with general spiral-like appearance in a system with differential rotation. The central problem is indeed to account for the more *permanent* features such as the spiral pattern over the whole disk. In this note, we outlined such a theory in terms of quasi-stationary self-sustained density waves. Such a gravitational theory has some obvious consequences. For example, pure circular motion is usually assumed in the reduction of data in 21-cm radio observations for the purpose of locating the spiral arm. The present theory would suggest the need to re-examine such data reduction since there are expected some systematic radial and circular motions of the gas distributed in the form of a spiral-like flow field.

† On leave of absence from Massachussetts Institute of Technology, Cambridge, Mass.

[1] Lin, C. C., and F. H. Shu, *Astrophys. J.*, **140**, 646–655 (1964).

[2] B. Lindblad [*Stockholm Obs. Ann.*, **22**, 3–20 (1963)] arrived at the same hypothesis from somewhat different considerations, with emphasis on individual orbits (theory of dispersion orbits). The present theory is based on statistical considerations, but is doubtless related to his theory. For example, the relations (4.1) and (4.2) are obviously related to Lindblad resonance which occurs at locations where $\kappa^2 - (\omega - m\Omega)^2 = 0$. There are also significant differences. Thus, in the absence of the gas, the condition for resonance is obtained from (4.1) for a *finite* extent in the galactic disk, as suggested by B. Lindblad, only if we set the reduction factor $\mathfrak{F}_\nu(x)$ equal to zero, i.e., if we neglect the spiral gravitational field due to the "cooperative" effect of the stars. See Paper I and later communications for further comments on the theories of B. Lindblad and P. O. Lindblad.

[3] The reduction factor applied to the gas may be taken either from the stellar dynamical theory or the gas dynamical theory. The results are substantially the same. Curves of similar nature are obtained when the mean square turbulent velocity is changed roughly in proportion to the amount of gas present.

[4] Toomre, A., *Astrophys. J.*, **139**, 1217 (1964).

[5] Kalnajs, A. J., Ph.D. thesis, Department of Astronomy, Harvard University (1965).

[6] The value $x = 7$ chosen is based on marginal stability in a pure stellar sheet for the most unstable waves in the sense of Toomre. If we make use of Fig. 2 to allow for the effect of the gas, a slightly higher value of x (perhaps $x = 8$) should be chosen. This does not modify the conclusions to any significant extent,